

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 16 日現在

機関番号：62615

研究種目：基盤研究(B) (一般)

研究期間：2014～2016

課題番号：26280066

研究課題名(和文) 声の生体検知を用いたセキュアな話者照合システムの実現

研究課題名(英文) Investigation on secure speaker verification based on vocal liveness detection

研究代表者

山岸 順一 (Yamagishi, Junichi)

国立情報学研究所・コンテンツ科学研究系・准教授

研究者番号：70709352

交付決定額(研究期間全体)：(直接経費) 12,500,000円

研究成果の概要(和文)：生体認証の中でも、マイクがあれば気軽に利用可能である音声による生体認証(話者照合)が注目を浴びており、社会への導入も進んだ。しかし近年飛躍的に向上した音声合成および声質変換技術を悪用することで話者照合を詐称できることも報告されている。そこで、本研究では、認証対象が真に生きている人間であるかどうかを判断する生体検知技術を話者照合に先駆的に導入した。大規模詐称音声コーパスを世界に先駆け構築し、話者照合システムの生体検知精度を競うイベントも開催した。さらにポップノイズという呼気で音声波形が歪む現象を利用し、音声が生体検知システムに発声したものなのか、スピーカー等で再生されたものなのかを識別できることも示した。

研究成果の概要(英文)：Among various biometrics, speech-based biometrics called speaker verification becomes popular recently and has been adopted for various applications in society and industry because it can be conveniently utilized based on a microphone only. However, it is also reported that spoofing of speaker verification is possible using recently advanced speech synthesis and voice transformation techniques. For this problem, we have introduced a novel technique called liveness detection or anti-spoofing for speaker verification systems to verify if an inputted speech signal is really produced by a live human or not. We have also constructed the first large-scale spoofed speech database in the world and have organized events where participants compete the accuracy of their anti-spoofing techniques on the same corpus. Moreover, we showed that it is possible to discriminate live human speech from speech replayed from loudspeakers using a pop noise, which is a distorted signal caused by breath.

研究分野：音声情報処理

キーワード：生体認証 話者照合 詐称 生体検知 音声合成 声質変換

1. 研究開始当初の背景

ユビキタス社会の到来により人間社会にとって、建物や携帯電話、パソコンなどのセキュリティが非常に重要になっており、近年、指紋や静脈など様々な生体情報を用いた生体認証が普及して始めている。しかし同時に、生体認証システムに対する詐称技術も先端ICT技術により高度化しており、コンピュータを利用した詐称の検出アルゴリズムの開発が急務になっている。ヨーロッパにおいても、EU FP7 プロジェクト TABULA RASA (<http://www.tabularasa-euproject.org/>) が中心となって、生体認証システムに対する詐称検出の検討を行うなど、国内外を問わず非常に活発に研究が行われていた。

生体認証の中でも、マイクロホンが一つあれば気軽に使用が可能であるという点から、音声を用いた生体認証、話者照合が注目を浴びており、社会への導入も進んでいた。

その一方、機械に発話させる技術「音声合成」、および声を変換する技術「声質変換」も飛躍的に向上し、少量の発話データがあれば個人の話者性を反映させた自然性の高い合成音声を作成することが可能となってきた。この声のクローン技術は音声障害者の会話支援技術として非常にニーズが高いものの、話者照合においては、クローンされた合成音声を用いた詐称が問題となってくる。実際、従来の話者照合システムで合成音声と実音声を識別することは難しく、合成音声を用いた詐称に対する脆弱性が数多く報告されていた。

2. 研究の目的

研究代表者は、位相スペクトルを使った詐称検出法を提案し、約90%の精度で人間の音声と合成音声を識別可能であることを示してきた。しかしながら、この検出法は、現在の音声合成・声質変換メカニズムを暗に仮定した対処法であり、クローン音声による詐称問題に対するセキュアな解決法とは言えない。抜本的な問題解決のためには、今後如何に音声合成や声質変換の性能が向上しようと頑健に詐称を検出できるアルゴリズムを研究、開発する必要がある。他のバイオメトリクスにも同様の問題が存在し、解決策の1つとして「生体検知」が着目されている。

生体検知とは、認証している対象が本当に生きている人間の一部であるかどうかを判断するプロセスのことであるが、話者照合システムでは生体検知という概念はこれまで全く導入されていない。そこで、本研究では、「声の生体検知」メカニズムを先駆的に導入したセキュアな話者認証システムを構築することを目的とする。

3. 研究の方法

(1) 話者認識と音声合成・声質変換は、これまで異なる分野であり、それぞれ独立して研究がなされてきた。合成音声を用いた詐称に対する話者照合システムの脆弱性が近年数多く報告されているものの、これ

まで、話者認識と音声合成分野を横断して使用可能なコーパスは存在しなかった。そこで、研究の第一段階では研究土壌の構築として、詐称者の視点から録音や音声合成、声質変換など現在考え得るすべての攻撃法を網羅した詐称者コーパスを作り上げる。このコーパスを作成することで、これまで異分野として研究されていた話者認識と音声合成を分野横断的に研究することが可能になり、詐称検出法をも同一のコーパスで統一的に評価することが可能になる。またこのコーパスを一般公開およびコーパスを利用した国際的なコンペティション「ASVSpooof Challenge」を開催することで、声の詐称検出を学術・研究分野として認知、確立させるとともに、先導的役割を果たす。

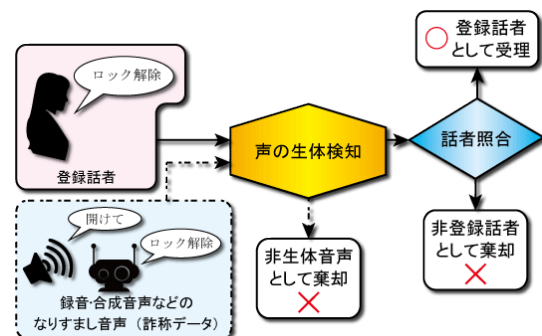


図1 システムの全体図

(2) これまでの声の詐称検出方法は音声信号における自然音声と合成音声の違いの差を見つけ出すものであった。しかし、詐称を真に見抜くためには差に着目するのではなく、抜本的な解決法が必要である。そこで本研究では、これとは原理的に異なり、認証対象の入力音声が入口の口から実際に発声されているかどうかを判定する方法を検討する。具体的には、スピーカーやロボットの口などの音声再生装置からは起こりえないが、人間の発声においては必ず起こる現象を捉え、これを声の生体情報として検知する。

そこで、破裂音等に伴う呼気により起こされた特殊なノイズ「ポップノイズ」をあえて収録し、直流に近い低周波成分を比較することを生体検知として行う。このポップノイズが適切な箇所にあると判断できれば、呼気を伴う音声、つまり生きた人間の口から実際に発声された音声と判断できる。

(3) 提案生体検知手法は、話者認証とは独立したモジュールであり、既存の話者照合システムを改変することなく利用することが出来る。SVM や i-Vector といった最先端の話者照合システムに対しても詐称可能であることも報告されていることから、これらの従来システムに声の生体検知を導入し、その有効性を確かめる。また同時に生体認証の普及を目指し、話者照合システム自身の識別性能の一層改善も行う。生体検知および識別モデルの二つの観点から検討&改善を行う事で、安全安心な生体認証の早期実現に向けて研究を邁進させる。

4. 研究成果

(1) 話者照合に対する攻撃は、収録音声のプレイバック、音声合成、声質変換などで行われることが知られていたが、それぞれのリスクを定量的に評価できるコーパスはこれまで存在しなかった。そこで、研究代表者および国内外の共同研究者が種々の音声合成と声質変換技術を利用し、大規模詐称音声コーパスを2014年に世界に先駆け構築した。2015年にはそのコーパスを世界の16研究機関及び企業に公開することで、話者照合システムの生体検知精度を競うASVspoof Challenge 2015を開催した。その後、プレイバック攻撃による詐称音声を集めたコーパスも新たに作成し、2017年に49研究機関及び企業に公開、ASVspoof Challenge 2017を開催した。

この大規模データベースの公開と二回のASVspoof Challengeにより、音声合成と声質変換攻撃に対する生体検知精度は約99%まで一気に向上し、また、これまで難関と思われていたプレイバック攻撃に対する生体検知精度も約94%の精度まで向上した。その他、ASVspoof challenge 2015では、学習データに含まれている5種の既知の合成方法は精度良く行えるのに対し、学習データに含まれていない5種の未知の音声合成方法や声質変換方法の検知は、既知の方法に比べ検知が難しく、エラー率が4倍ほど高いと言う問題を明らかにした。ASVspoof challenge 2017では、プレイバック攻撃に利用されるマイク、プレイバックの際に音声を再生するスピーカー、および収録環境を要因として考えた112種のプレイバックデータを既知・未知のデータに分け、検知が容易である条件、および、検知が難しい条件を明らかにした。

さらに、ASVspoof Challengeでは詐称音声の構築作業と生体検知技術の開発は全く別々の研究グループが独立して行った。具体的には、詐称音声の構築は音声合成・声質変換コミュニティーが行い、生体検知の開発は話者照合コミュニティーが行った。これにより、詐称音声を構築する側と話者照合システムを構築する側がともに事前知識を持たない形となり、より公正で現実に即した生体検知技術の評価を行えるようになった。

(2) 入力音声が実際に人間から発声されたものなのか、スピーカー等で再生されたものなのかを識別する枠組みとして声の生体検知法を世界にさきがけて提案した。この声の生体検知法ではポップノイズという話者の呼気等で音声波形が歪む現象を検出し利用した。

ポップノイズが発生すると音声認識や話者照合の性能低下が起こる可能性があるため、通常はポップノイズが発生することを、ポップフィルタ等により回避する。本手法では、あえてポップフィルタを外し、内部に呼気の影響を直接受けさせるようにしたマイクと通常のマイクの2チャンネルにより同

時収録し、この2チャンネル間の周波数上の差分を生体検知の特徴量とすることで、生体音声と再生音声の識別を行えることを示した。具体的には、17名の話者を上記2チャンネルヘッドセットマイクにより収録し、実験を行なった結果、本手法による生体検知精度は94%(EER=5.88%)であるという結果になり、ポップノイズ検知アルゴリズムを用いることで話者照合の脆弱性を補うことができることがわかった。

(3) 提案したポップノイズ検出法による声の生体検知手法は話者照合システムの前段の処理として用いることで話者照合システムに入力されるなりすまし攻撃を大幅に減らすことができる。話者照合システムの本来のエラー率が4.28%であってもなりすまし攻撃を混ぜるとエラー率が6.07%と悪化してしまう。生体検知と話者照合を組み合わせることでこのエラー率を4.4%とほぼなりすまし攻撃がない状態のシステム性能まで引き上げられることが分かった。また、ポップフィルタを含まない音声は背景雑音が入りやすいという特徴があるため、耐雑音性の高い話者照合システムを用いる必要がある。これまでに、背景雑音のSNRが異なる音声を用いて学習したUBM-GMMと呼ばれる話者照合システムを統合して用いることで雑音に対する頑健性が向上することを報告した。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 3 件)

Zhizheng Wu, Junichi Yamagishi, Tomi Kinnunen, Cemal Hanilci, Md Sahidullah, Aleksandr Sizov, Nicholas Evans, Massimiliano Todisco, Hector Delgado, "ASVspoof: the Automatic Speaker Verification Spoofing and Countermeasures Challenge", Special Issue on Spoofing and Countermeasures for Automatic Speaker Verification, IEEE Journal of Selected Topics in Signal Processing, vol.11, issue 4, pp. 588 - 604, 2017年6月, 査読有り, <https://doi.org/10.1109/JSTSP.2017.2671435>

Zhizheng Wu, Phillip L. De Leon, Cenk Demiroglu, Ali Khodabakhsh, Simon King, Zhen-Hua Ling, Daisuke Saito, Bryan Stewart, Tomoki Toda, Mirjam Wester, and Junichi Yamagishi, "Anti-Spoofing for Text-Independent Speaker Verification: An Initial Database, Comparison of Countermeasures, and Human Performance", IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol.24, issue 4, pp.768-783, 2016年4月, 査読有り, <https://doi.org/10.1109/JSTSP.2017.2671435>

〔学会発表〕(計 32 件)

Ryosuke NAKANISHI, Sayaka SHIOTA, and Hitoshi KIYA, "Ensemble Based Speaker Verification Using Adapted Score Fusion in Noisy Reverberant Environments," Proc. APSIPA Annual Summit and Conference, Jeju Korea, 2016 年 12 月 13 日-16 日

Sayaka SHIOTA, Fernando Villavicencio, Junichi YAMAGISHI, Nobutaka ONO, Isao ECHIZEN, and Tomoko MATSUI, "VOICE LIVENESS DETECTION FOR SPEAKER VERIFICATION BASED ON A TANDEM SINGLE/DOUBLE-CHANNEL POP NOISE DETECTOR," Proc. The Speaker and Language Recognition Workshop Odyssey, pp.259-263, Spain, Bilbao, 2016 年 6 月 21 日-24 日

Mirjam Wester, Zhizheng Wu, Junichi Yamagishi, "Human vs Machine Spoofing Detection on Wideband and Narrowband Data", Interspeech 2015, pp.2047-2051, Dresden, Germany, 2015 年 9 月 6 日-10 日

Sayaka Shiota, Fernando Villavicencio, Junichi Yamagishi, Nobutaka Ono, Isao Echizen, Tomoko Matsui, "Voice liveness detection algorithms based on pop noise caused by human breath for automatic speaker verification", Interspeech 2015, pp.239-243, Dresden, Germany, 2015 年 9 月 6 日-10 日

Zhizheng Wu, Tomi Kinnunen, Nicholas Evans, Junichi Yamagishi, Cemal Hanile, Md Sahidullah Aleksandr Sizov, "ASVspoof 2015: the First Automatic Speaker Verification Spoofing and Countermeasures Challenge", Interspeech 2015, pp.2037-2041, Dresden, Germany, 2015 年 9 月 6 日-10 日

Zhizheng Wu, Ali Khodabakhsh, Cenk Demiroglu, Junichi Yamagishi, Daisuke Saito, Tomoki Toda, Simon King, "SAS: A SPEAKER VERIFICATION SPOOFING DATABASE CONTAINING DIVERSE ATTACKS", ICASSP 2015, pp.4440-4444, Brisbane, Australia, 2015 年 4 月 19 日-24 日

〔図書〕(計 3 件)

Nicholas Evans, Tomi Kinnunen, Junichi Yamagishi, Zhizheng Wu, Federico Alegre and Phillip de Leon, "Voice anti-spoofing" a book chapter in Handbook of Biometric Anti-Spoofing Sébastien Marcel, Mark S.Nixon, and Stan Z. Li (eds). Springer, 2014 年 8 月 ISBN:978-1447165231

〔産業財産権〕

出願状況 (計 1 件)

名称: 生体検知装置、生体検知法およびプログラム

発明者: 山岸順一、越前功、小野順貴、松井知子、塩田さやか

権利者: 同上

種類: 特許

番号: 特願 2014-166271

出願年月日: 2014 年 8 月 19 日

国内外の別: 国内

取得状況 (計 0 件)

〔その他〕

ホームページ等

Spoofing and Anti-Spoofing (SAS) corpus

<http://dx.doi.org/10.7488/ds/252>

Automatic Speaker Verification Spoofing and Countermeasures Challenge (ASVspoof 2015) Database

<http://dx.doi.org/10.7488/ds/298>

6. 研究組織

(1) 研究代表者

山岸順一 (YAMAGISHI, Junichi)

国立情報学研究所・コンテンツ科学研究系・准教授

研究者番号: 70709352

(2) 研究分担者

松井知子 (MATSUI, Tomoko)

統計数理研究所・モデリング研究系・統計的機械学習研究センター・サービス科学研究センター・教授

研究者番号: 10370090

越前功 (ECHIZEN, Isao)

国立情報学研究所・情報社会相関研究系・教授

研究者番号: 30462188

小野順貴 (ONO, Nobutaka)

国立情報学研究所・情報学プリンシプル研究系・准教授

研究者番号: 80334259

塩田さやか (SHIOTA, Sayaka)

首都大学東京・システムデザイン研究系・助教

研究者番号: 90705039

(3) 連携研究者

該当なし

(4) 研究協力者

該当なし